

# Forecasting and Risk

## (BANA 4090)

### Practical Time Series Analysis

Zhaohu(Jonathan) Fan

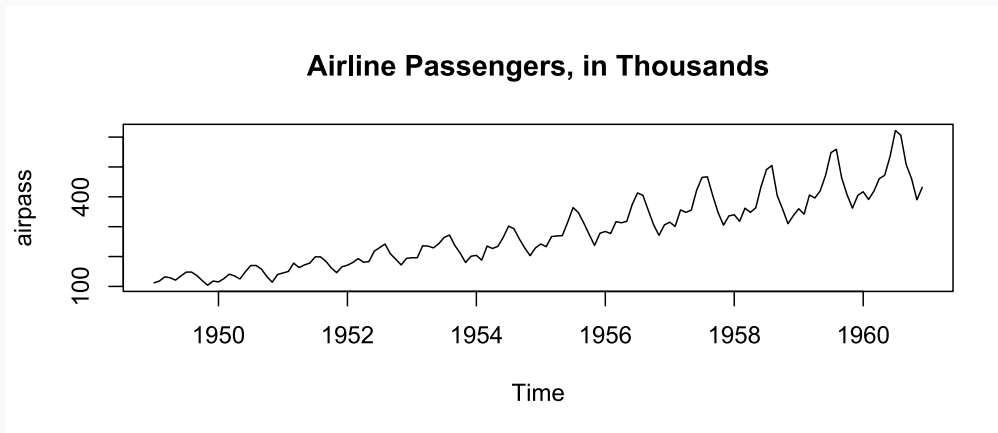
- Topics covered:
  - Chapter 1
    - Visualizing Time Series Data in R
  - Chapter 2

# Chapter 1

# Time series data

- Time series data form an ordered sequence of numbers, corresponding to an object like quantities, prices, counts, observed at or over a particular point **in time**.
- The "Airline Passengers" data set is an example of **time series data**.

```
plot(airpass)  
title(main="Airline Passengers, in Thousands")
```



# Time series patterns

- Describing a time series: trend, seasonality, cycles, changing variance, unusual features.
  - **Trend**: pattern exists when there is a long-term increase or decrease in the data.
  - **Seasonal** : pattern exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week).
  - **Cyclic** : pattern exists when data exhibit rises and falls that are *not of fixed period* (duration usually of at least 2 years).

# Seasonal or cyclic?

## **Differences between seasonal and cyclic patterns:**

- seasonal pattern constant length; cyclic pattern variable length
- average length of cycle longer than length of seasonal pattern
- magnitude of cycle more variable than magnitude of seasonal pattern

**The timing of peaks and troughs is predictable with seasonal data, but unpredictable in the long term with cyclic data.**

# Visualizing time series data in R

- Visualization is good practice to be able to understand the properties of the data.
  - Most time series coming from official data sources provide recordings at a regularly spaced set of such time points, such as every day, week, month, quarter, or year; this interval is called the **frequency** of the time series.
- A time series is stored in a `ts` object in R:
  - `ts` objects and `ts` function

For observations that are more frequent than once per year, add a `frequency` argument.

E.g., monthly data stored as a numerical vector `z`:

```
y ← ts(z, frequency=12, start=c(2003, 1))
```

# Chapter 2



# Correlation

- Correlation: measure of linear relationships (**a measure of the direction and strength of the relationship between two variables**)
- We use Pearson's  $r$  as a measure of the linear relationship between two quantitative variables. In a sample, we use the symbol  $r$ . In a population, we use the Greek letter ("rho"). Pearson's  $r$  can easily be computed using R.
- The correlation coefficient,  $\rho$ , measures linear relationships: Ranges over  $[-1, +1]$ 
  - A value of  $+1$  indicates a perfect positive (upward sloping) linear relationship between the two variables.
  - A value of  $-1$  indicates a perfect negative (downward sloping) linear relationship between the two variables.
  - A value of zero indicates no linear relationship between the two variables.

# Interpret correlation coefficient

- Correlation coefficient is comprised between -1 and 1:
  - -0.86 indicates a strong negative correlation : this means that every time  $x$  increases,  $y$  decreases.
  - 0 means that there is no association between the two variables ( $x$  and  $y$ ).
  - 0.87 indicates a strong positive correlation: this means that  $y$  increases with  $x$ .

Note: The closer  $r$  is to 0 the weaker the relationship and the closer to +1 or -1 the stronger the relationship (e.g.,  $r=-0.98$  is a stronger relationship than  $r=+0.78$  ); the sign of the correlation provides direction only.

# Interpret correlation coefficient

- To determine whether the correlation between variables is significant, compare the p-value to your significance level. Usually, a significance level (denoted as  $\alpha$  or alpha) of 0.05 works well.
- An  $\alpha$  of 0.05 indicates that the risk of concluding that a correlation exists—when, actually, no correlation exists—is 5%. The p-value tells you whether the correlation coefficient is significantly different from 0. (**A coefficient of 0 indicates that there is no linear relationship.**)
  - **P-value  $\leq \alpha$ :** The correlation is statistically significant. If the p-value is less than or equal to the significance level, then you can conclude that the correlation is different from 0.
  - **P-value  $> \alpha$ :** The correlation is not statistically significant. If the p-value is greater than the significance level, then you cannot conclude that the correlation is different from 0.

Questions? 😊